

采用欧式形态距离的负荷曲线近邻传播聚类方法

党倩<sup>1</sup>, 崔阿军<sup>2</sup>, 尚闻博<sup>1</sup>, 杨波<sup>2</sup>, 卫祥<sup>1</sup>

(1. 国网甘肃省电力公司信息通信公司, 730050, 兰州; 2. 国网甘肃省电力公司, 730000, 兰州)

**摘要:** 为提取电力负荷数据的有效信息,改善传统聚类方法在电力负荷数据中相似度衡量方式单一及聚类效果较差的问题,提出一种采用欧式形态距离的负荷曲线近邻传播(AP)聚类方法。使用五分位法将用电负荷曲线重表达为曲线形态变化特征序列,使用改进最长公共子序列算法衡量不同特征序列之间的模式匹配度,以此作为曲线之间的差异度;构造一种兼顾曲线整体分布特征和曲线形态变化特征的双尺度相似性度量方法,使用熵权法对两种特征进行自适应配比;将所提相似度衡量方法应用到 AP 聚类方法中,改进相似度矩阵计算方法,对用户典型日用电负荷曲线进行聚类。在标准合成时间序列数据集上进行了实验对比,结果表明:欧式形态距离度量方法能够有效区分负荷曲线的变化特征;所提方法具有较高的聚类质量和稳健性,相比其他相似度量方法,调整兰德系数提高了 9.0%~43.8%,DB 指标与标准集相差 0.014 3,在电力实测数据集上能对用户进行合理划分。

**关键词:** 智能电网;双尺度相似性度量;曲线形态变化特征;负荷聚类;五分位法

**中图分类号:** TM714 **文献标志码:** A

**DOI:** 10.7652/xjtuxb202201019 **文章编号:** 0253-987X(2022)01-0165-12



OSID 码

Affinity Propagation Clustering Method of Typical Load Curve with Euclidean Morphological Distance

DANG Qian<sup>1</sup>, CUI Ajun<sup>2</sup>, SHANG Wenbo<sup>1</sup>, YANG Bo<sup>2</sup>, WEI Xiang<sup>1</sup>

(1. State Grid Gansu Information & Telecommunications Company, Lanzhou 730050, China;

2. State Grid Gansu Electric Power Corporation, Lanzhou 730000, China)

**Abstract:** An improved affinity propagation clustering algorithm with Euclidean morphological distance was proposed to extract effective information from load curve and improve the poor clustering effect of the traditional clustering algorithms. The quantile method was used to extract and re-express morphological characteristic of the original load curve, the improved longest common subsequence algorithm was introduced to measure the pattern matching distance between different user load consumption behavior sequences, and the measurement was taken as the degree of difference between the curves. Then a dual-scale similarity measurement considering the overall distribution of the curve and the change pattern of the curve shape was conducted, and the entropy weight method was used to adaptively match the weight between the two characteristics. The proposed similarity measurement was applied to the affinity propagation clustering algorithm to improve the similarity matrix calculation scheme, and the typical daily load curve of users was clustered. Taking the standard synthetic time series data as an example, clustering results

indicate that the Euclidean morphological distance measurement method can effectively distinguish the dynamic change characteristics of load curves. The proposed method has higher clustering quality and algorithm robustness, increases the adjusted Rand index from 9.0% to 43.8%, differs from the standard set in DB index by 0.014 3, and can distinguish users on the measured data set appropriately.

**Keywords:** smart grid; dual-scale similarity measurement; curve morphological characteristic; load clustering; quintile method

近年来,中国智能电网的建设大力推进,智能电表的大量普及使用户用电数据的收集变得简单易行,更细粒度的用电负荷数据保证了供需交互的稳定性<sup>[1]</sup>。然而,海量信息的涌入意味着电力数据信息的冗余化与杂乱化。因此,如何运用数据挖掘技术,从用电数据中提取有效信息,分析用户用电模式,为电力决策提供参考,已被认为是电网领域的重要研究内容<sup>[2]</sup>。

电力负荷曲线聚类是电力数据挖掘中的常见技术,其目的是提取用户用电负荷曲线的分布特征,找出其中用户行为的相似点,进而归类分析用户用电模式<sup>[3]</sup>。目前,各类聚类分析技术方法已被应用于电力负荷聚类中,包含基于划分的聚类方法<sup>[4-5]</sup>、层次聚类方法<sup>[6]</sup>、基于图的聚类方法<sup>[7]</sup>等。

相似度量方法决定了聚类方法对于负荷曲线变化特征的敏感度,是影响聚类质量的重要指标<sup>[8]</sup>。主流方法大多关注于两条曲线相同采样点的负荷差异,最常见的方法为欧氏距离,通过计算相同采样点之间的距离均值,判断曲线相似程度。为了改善聚类质量:卜凡鹏等采用双层聚类的方式,外层结合皮尔森系数,内层使用欧氏距离,能够更细粒度地控制聚类结果<sup>[9]</sup>;徐胜蓝等将欧氏距离与余弦相似距离结合,提出双尺度相似谱聚类集成方法,得到较好的效果<sup>[10]</sup>;冯志颖等引入推土机距离,结合欧氏距离从横纵向两方面分析负荷曲线,可用于检测异常用电<sup>[11]</sup>;李阳等使用差分算法及分位数提取负荷曲线斜率变化特征,根据对应时间点特征值是否相同进行归类,忽略了负荷曲线的位移变化特征<sup>[12]</sup>。虽然部分方法将两种相似度量方法行结合,实现不同方法优势的互补,但这类相同采样点之间点对点匹配的方法只能观察到负荷曲线之间整体的差异性,而忽略了用户用电的时间差异,随着负荷采样频率的提高,这类方法所能表达的信息将更加有限。针对这些问题,Teeraratkul 等将动态时间弯曲距离(DTW)引入负荷聚类,实现了对曲线的动态匹配,但未能解决方法时间开销过大的问题<sup>[13]</sup>。

目前负荷聚类的度量方法主要有以下不足:①多数方法仅考虑负荷曲线相同采样点的数值差异,忽略了曲线的位移变化;②现有动态匹配方法计算开销过大,可能导致过度弯曲。

针对现有研究的局限性,本文提出了一种采用欧式形态距离的负荷曲线近邻传播(AP)聚类方法 EMD-AP。将用电负荷曲线重表达为离散曲线形态变化特征,通过基于模式匹配距离的最长公共子序列(LCS)算法衡量不同特征序列的差异性,结合欧氏距离构造兼顾负荷曲线动态变化特征以及整体分布特征的相似度度量方法,使用熵权法对这2种距离方法进行权重赋值,采用 AP 聚类对不同用户负荷曲线进行区分。本文方法具有时间开销低和聚类准确性高等优点。

## 1 曲线相似性度量

### 1.1 负荷曲线特性分析

为了有效分辨不同负荷曲线之间的差异,本文从负荷曲线的整体分布特征和形态变化特征这2种特征来衡量负荷曲线之间的相似性。整体分布特征对应曲线相同时刻采样点之间的负荷差异度,反映了负荷曲线的整体相似度。形态变化特征对应曲线在采样时间内整体的形状变化特征的匹配程度,反映了形态变化情况。

### 1.2 曲线整体分布特征

欧氏距离是常见的点对点匹配方式,本文采用欧氏距离来度量负荷曲线的整体分布特征。给定2个等长用电负荷曲线  $X = [x_1, x_2, \dots, x_n]$  和  $Y = [y_1, y_2, \dots, y_n]$ ,  $X$  和  $Y$  之间的欧氏距离定义为

$$D_{ed}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

### 1.3 曲线形态变化特征

1.3.1 曲线形态变化特征提取 用电负荷曲线反映了用户在不同时刻的用电量,其实质是用户一系列用电行为的叠加。因此,本小节将用户用电负荷曲线转换为描述用户在不同时刻用电行为的特征序

列。为了度量不同用户用电行为特征的差异性(同时忽略不同用户用电量级存在的差异),本文基于不同用户在每一时刻的用电量与用电最小负荷的差值,并且结合分位数对其用电特征进行提取与统一性重表达,得到使用离散类属性表达的曲线形态变化特征序列。

首先,对于用电负荷曲线  $X=[x_1, x_2, \dots, x_n]$ , 获取其最大值  $x_{\max}$  与最小值  $x_{\min}$ , 计算两者的差值记为  $x_{\text{range}}$ ; 然后, 采用分位数刻画其数据分布特征, 为了去除噪声扰动并保留更多有效信息<sup>[14]</sup>, 采用五分位数组合  $\tau=(0.1, 0.3, 0.5, 0.7, 0.9)$  计算曲线形态变化特征序列  $X_d=[x_{d1}, x_{d2}, \dots, x_{dn}]$ , 公式为

$$x_{di} = \begin{cases} 5, & d_{di} \geq Q_5 \\ 4, & Q_4 \leq d_{di} < Q_5 \\ 3, & Q_3 \leq d_{di} < Q_4 \\ 2, & Q_2 \leq d_{di} < Q_3 \\ 1, & Q_1 \leq d_{di} < Q_2 \\ 0, & d_{di} < Q_1 \end{cases} \quad (2)$$

式中:  $Q_1, \dots, Q_5$  为对应的五分位数值,  $Q_i = \tau_i x_{\text{range}} + x_{\min}$ ;  $d_{di}$  为负荷曲线  $X$  在第  $i$  时刻的负荷  $x_i$  与曲线最小值  $x_{\min}$  的差值。

图 1 为曲线形态变化特征提取图。可以看出: 五分位法将负荷曲线转化为离散的形态变化特征序列, 完整保留了负荷曲线的波动特征, 同时消除了负荷曲线用电负荷量级之间的差异性。

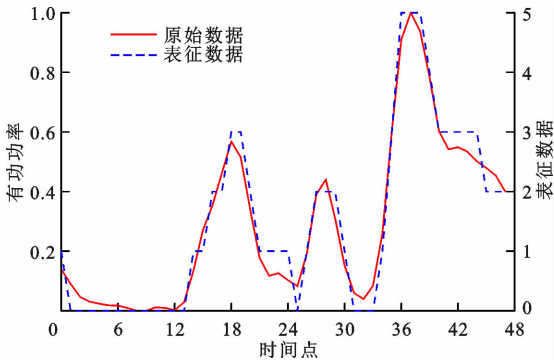


图 1 负荷曲线形态变化特征提取图

Fig. 1 Morphological feature extraction of load curves

**1.3.2 曲线形态变化特征度量** LCS 是一种基于互相关的距离度量方法, 用于寻找两条时序曲线的最长公共子序列, 该方法能够抵抗一定程度的环境噪声和数据失真的情况, 常用于模式匹配<sup>[15]</sup>。本文通过特征提取的方式将原始用户每一时刻的用电数据统一性重表达为仅由 6 种离散型类属性组成的曲线形态变化特征序列, 消除了原始数值型数据之间

存在的量级间差异。因此, 可以提取出不同特征序列之间的公共子序列, 并基于序列的长度来度量用户用电行为之间的相似性。

为了最大程度保留曲线的有效信息, 本文对曲线进行等时间分辨率的形态重表达, 使用模式匹配距离衡量用电行为特征的差异性。

模式匹配距离计算式为

$$d_{\text{mode}}(x, y) = \begin{cases} 0, & x = y \\ 1, & x \neq y \end{cases} \quad (3)$$

对于给定的特征序列  $X=[x_1, x_2, \dots, x_m]$  和  $Y=[y_1, y_2, \dots, y_n]$ , 构建距离矩阵  $D \in \mathbf{R}^{m \times n}$ , 矩阵元素  $d_{i,j}$  代表序列  $X$  的子序列  $X_i$  与序列  $Y$  的子序列  $Y_j$  最大公共子序列的长度。采用动态规划算法求解其最长公共子序列, 动态规划算法通常以从部分到整体的顺序进行解决, 后一状态的结果往往由前一状态决定。对于矩阵中任意元素  $d_{i,j}$  需满足: ①当  $d_{\text{mode}}(x_i, y_j) = 0$  时, 公共子序列长度即为  $d_{i,j} = d_{i-1, j-1} + 1$ ; ②当  $d_{\text{mode}}(x_i, y_j) = 1$  时,  $d_{i,j}$  为  $d_{i-1, j}$  和  $d_{i, j-1}$  的最大值; ③为合理限制两条负荷曲线特征序列之间的匹配关系, 使用阈值参数  $\epsilon$  控制序列之间参数匹配的窗口尺寸, 即  $d_{i,j}$  中下标  $i$  和  $j$  的距离需要小于  $\epsilon$ 。最终, 两条特征序列之间的最长公共子序列长度即为  $d_{n,m}$ 。

改进 LCS 算法计算方式为

$$f_{\text{LCS}}(X, Y) = \begin{cases} 0, & m = 0 \text{ or } n = 0 \\ 1 + f_{\text{LCS}}(X_{i-1}, Y_{j-1}), & d_{\text{mode}}(x_i, y_j) = 0 \text{ and } |j - i| \leq \epsilon \\ \max(f_{\text{LCS}}(X_{i-1}, Y_j), f_{\text{LCS}}(X_i, Y_{j-1})), & \text{others} \end{cases} \quad (4)$$

式中:  $m$  和  $n$  分别代表两条负荷曲线  $X$  和  $Y$  的长度, 本文实验中  $m=n$ ; 阈值参数  $\epsilon$  的取值范围为  $[0, m]$ 。当  $\epsilon=0$  时, LCS 算法退化为对两条曲线对应时间点的元素进行比较, 算法时间复杂度为  $O(n)$ ; 当  $\epsilon=m$  时, LCS 算法为基于动态规划的 LCS 算法, 算法时间复杂度为  $O(n^2)$ 。本文数据集中, 每天包含 48 个采样点, 假设同一类用电用户对于电气设备的使用最多存在 2 h 的偏差<sup>[13]</sup>, 即本文中阈值参数  $\epsilon$  设为 4。

不同时间序列之间的形态距离为

$$D_{\text{md}}(X, Y) = \text{len}(X_d) - f_{\text{LCS}}(X_d, Y_d) \quad (5)$$

式中:  $D_{\text{md}}$  表示两条用电行为序列之间的形态距离;  $\text{len}(X_d)$  为曲线形态特征序列  $X_d$  的长度, 即  $D_{\text{md}}$  取值范围为  $[0, \text{len}(X_d)]$ 。  $D_{\text{md}}$  越小说明 2 条时序曲线的公共特征子序列越长, 曲线的形态变化特征越

相似。

## 2 采用欧式形态距离的 AP 聚类方法

### 2.1 引入形态距离的必要性

对于负荷曲线,在衡量其相似性时需要考虑到位移变化、尺度变化和噪声变化特性,当负荷曲线发生这些变化后,相似度保持不变<sup>[16]</sup>。

传统采用欧氏距离的相似度量方法,仅考虑了两条负荷曲线相同时刻点之间的用电负荷差异,将导致聚类方法容易对具有位移变化的负荷曲线进行误匹配,造成聚类偏差。DTW 能够对序列进行延展或压缩,达到识别曲线位移变化的目的。但是,这类方法通常是基于两点之间的欧式距离,对于用电模式相同但是进行了一定尺度变化的用户也很难做到精确分类,而且时间开销大。

由 1.3 小节可知,本文形态距离度量方法的特征在于:①相比欧氏距离,形态距离通过动态规划的方法,可以实现具有位移变化的曲线元素之间的匹配,保证了对曲线整体动态特性的学习,适用于 2 条曲线之间存在一定的时移偏差或者两条曲线负荷峰值窗口大小不一致的情形;②相比 DTW,形态距离根据负荷曲线的负荷波动提取原始曲线的形态变化特征,在进行特征匹配的时候使用模式匹配距离,能有效忽略曲线的尺度变化,对曲线的负荷值不敏感,更适用于 2 条曲线形态相似但距离相近的情形。同时,在计算成本上,本文所提方法与 DTW 方法在不考虑阈值参数设定的情况下,时间复杂度均为  $O(n^2)$ ,但是形态距离将曲线数值数据转换为定性数据,使用异同的思想衡量曲线形态差异,每次匹配只需要一次判等操作,计算过程简单,具体执行时间远小于 DTW 的。图 2 为形态相近但有位移变化的负荷曲线。

由图 2 可以看出:3 条曲线均属于双峰型负荷

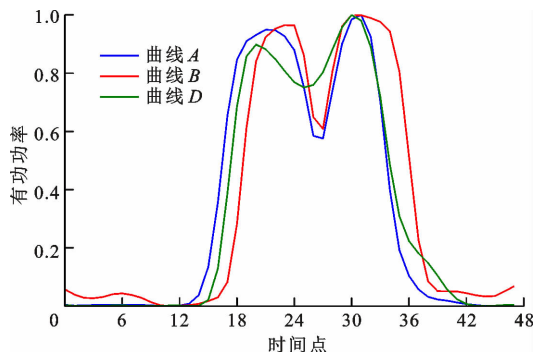


图 2 负荷曲线形态特征对比

Fig. 2 Morphological feature comparison of load curves

曲线,负荷曲线 A 和 B 形态更加相近,但是曲线 A 相对于曲线 B 有 3 个采样点的位移变化,曲线 A 和 D 距离更近但是形态差异更大。分别使用 3 种相似度量方式计算 3 类曲线的相似性。首先,采用欧氏距离计算,得到  $D_{ed}(A,B)=1.368\ 9$ ,  $D_{ed}(A,D)=0.622\ 2$ 。可以看出,如果仅使用欧氏距离作为曲线相似度的度量方式,那么对于这类出现位移变化的曲线识别效果较差。接着,采用 DTW 距离计算,得到  $D_{dtw}(A,B)=0.083\ 1$ ,  $D_{dtw}(A,D)=0.102\ 4$ 。采用形态距离进行计算,得到  $D_{md}(A,B)=5$ ,  $D_{md}(A,D)=10$ 。可以看出,DTW 方法和形态距离方法均能得到曲线 B 与曲线 A 相似度更高的结论,更符合实际情况。

### 2.2 欧式形态距离描述

欧氏距离是一种基于点对点的曲线相似性度量方法,通过计算负荷曲线对应时间点之间的距离平均值获得曲线的相似度特征,能有效识别曲线的整体分布特征。形态距离方法通过对原始用电曲线进行离散属性表示,然后与基于 LCS 的动态特征匹配,得到曲线的特征匹配度,能有效度量曲线的形态变化特征以及相应的趋势匹配信息。

本文综合考虑曲线形态的整体分布特征和曲线的波动特征,构造一种负荷曲线双尺度相似性度量——欧氏形态距离,计算式为

$$D_{emd}(X,Y)=\alpha D_{md}(X,Y)+\beta r_{m,e} D_{ed}(X,Y) \quad (6)$$

式中: $D_{emd}(X,Y)$ 为负荷曲线 X 和 Y 之间的欧式形态距离; $\alpha$ 和 $\beta$ 分别为相似度矩阵的权重系数,为保证权重参数的客观性,使用改进熵权法<sup>[17]</sup>进行赋值; $r_{m,e}$ 为形态距离与欧氏距离的比值系数,考虑到不同度量方法取值范围的不同,使用比例系数 $r_{m,e}$ 实现两种方法取值范围的统一,计算式为

$$r_{m,e} = \frac{\max(D_{md})}{\max(D_{ed})} \quad (7)$$

### 2.3 聚类质量评价

2.3.1 内部评价指标 聚类内部评价指标一般用于对无标签数据聚类质量的评价,要求数据具有较小的类内聚合度和较大的类间差异度。本文采用在电力数据集有较好聚类评价效果的 DB 指标<sup>[18]</sup>。

DB 指标能综合计及簇内的集聚度和簇间的分离度,计算式为

$$I_{DB} = \frac{1}{K} \sum_{k=1}^K R_k \quad (8)$$

$$R_k = \max_{k \neq j} \frac{d(X_k) + d(X_j)}{d(C_k, C_j)} \quad (9)$$

式中: $d(X_k)$ 和 $d(X_j)$ 分别为簇 $C_k$ 和簇 $C_j$ 中样本点与样本中心的平均欧氏距离,表示一个簇内样本间的聚合度; $d(C_k, C_j)$ 为簇 $C_k$ 和簇 $C_j$ 中样本中心之间的平均欧氏距离,表示两个类簇之间的分离度。

考虑到传统基于欧氏距离的 $I_{DB}$ 已不能对本文方法进行准确评价,提出基于欧式形态距离的修正指标 MDB,计算式为

$$I_{MDB} = \frac{1}{K} \sum_{k=1}^K \max_{k \neq j} \frac{d_{emd}(X_k) + d_{emd}(X_j)}{d_{emd}(C_k, C_j)} \quad (10)$$

式中: $d_{emd}(X_k)$ 和 $d_{emd}(X_j)$ 为簇 $C_k$ 和簇 $C_j$ 中样本点与样本中心的平均欧氏形态距离; $d_{emd}(C_k, C_j)$ 为簇 $C_k$ 和簇 $C_j$ 样本中心之间的平均欧氏形态距离。

$I_{DB}$ 和 $I_{MDB}$ 都是将两簇之间的类内聚合度和类间分离度做比,因此数值越低代表聚类质量越高。在选择最优聚类数时一般寻找其极小值点。

**2.3.2 外部评价指标** 外部评价指标用于在样本真实标签已知的情况下,将样本真实标签及聚类结果进行对比,以评估聚类有效性。选择调整兰德系数(AR)指标和 FM 指标共同评价各方法聚类结果的有效性。

(1)AR 指标。AR 指标是一种常见的聚类外部评价指标,通过计算在真实标签和聚类结果中被分配在相同或不同类簇的样本对数来评价聚类有效性,计算式为

$$I_{AR} = \frac{I_{RI} - E(I_{RI})}{\max(I_{RI}) - E(I_{RI})} \quad (11)$$

$$I_{RI} = \frac{N_{TP} - N_{TN}}{C_m^2} \quad (12)$$

式中: $I_{RI}$ 为兰德指数; $E(I_{RI})$ 为兰德指数的期望值; $\max(I_{RI})$ 为兰德指数的最大值; $N_{TP}$ 为真实标签与聚类结果中均被归为同一类的样本对数量; $N_{TN}$ 为真实标签与聚类结果中均被归为不同类的样本对数量; $C_m^2$ 为从 $m$ 个样本里任取 2 个样本的组合数。

FM 指标为成对的准确率和召回率的几何平均值,具体定义为

$$I_{FM} = \frac{N_{TP}}{\sqrt{(N_{TP} + N_{FP})(N_{TP} + N_{FN})}} \quad (13)$$

式中: $N_{TP}$ 为在真实标签与聚类结果中均被归为同一类的样本对数量; $N_{FN}$ 为在预测标签组中不属于同一簇而属于真实标签组的样本对数量; $N_{FP}$ 为在预测标签组中属于同一簇而不属于真实标签组的样本对数量。

$I_{AR}$ 取值范围为 $[-1, 1]$ , $I_{FM}$ 取值范围为 $[0, 1]$ ,数值越大表示聚类结果越贴近真实情况,取值为 1 时表明聚类结果与真实标签一致。

## 2.4 近邻传播聚类方法介绍

AP 聚类方法认为所有样本都有成为聚类中心的可能性,它通过选举的方式选拔出一组高质量的样本作为聚类中心。相比于传统的聚类方法,AP 聚类方法通过控制参考度调节聚类数<sup>[19]</sup>,并且具有更高的稳定性<sup>[20]</sup>。本文引入欧式形态距离对负荷曲线相似度进行度量,提升聚类效果。

**2.4.1 相似度矩阵的构建** AP 聚类方法以成对样本之间相似度距离构成的相似性矩阵 $S$ 作为输入,矩阵元素 $S(i, j)$ 一般为序列 $X$ 和 $Y$ 之间欧氏距离的相反数。本文 $S(i, j) = -D_{emd}(X, Y)$ 。 $S(i, j)$ 越大则序列 $X$ 和 $Y$ 越相似。

除了相似度量方法的设定,AP 聚类方法还通过参考度 $p$ 和阻尼系数 $\lambda$ 控制方法的聚类效果。参考度 $p$ 对应为相似度矩阵的对角线元素 $S(i, i)$ ,可以通过调节参考度 $p$ 控制最终聚类结果的粒度,参考度越大则聚类数越少。本文 $p$ 的初始值取相似度矩阵的最小值,后续通过聚类评价指标进行调节。为了避免在聚类过程中出现参数振荡无法收敛的情况,本文将 $\lambda$ 的默认值设定为 0.9。

**2.4.2 AP 聚类方法实现过程**  $I_{DB}$ 和 $I_{MDB}$ 分别从整体分布情况以及局部动态变化情况两个方面描述聚类结果中各类簇的紧凑程度,数值越小代表各簇内负荷曲线分布越紧凑。本文综合考虑两种指标,以描述负荷曲线整体分布情况的 $I_{DB}$ 最小值对应聚类数作为最佳聚类数,并使用 $I_{MDB}$ 对其正确性进行验证。本文方法实现流程如图 3 所示。设定方法初始聚类数为 2,最大聚类数 $N_{max} = \sqrt{0.5N}$ , $N$ 为数据集中用电用户数量,通过调节 $p$ 得到不同的聚类数。本文方法具体步骤如下。

**步骤 1 数据预处理。**首先,对原始数据进行数据清洗,去除全 0、全天负荷不变以及缺失超出 10% 的负荷曲线,取 8 月工作日用电负荷曲线并取均值,作为典型日用电负荷曲线。为避免异常负荷值对重表达效果的影响,使用高斯滤波对曲线进行平滑处理,并进行极差标准化。

**步骤 2 参数初始化。**设定 AP 聚类的参考度 $p$ 、阻尼系数 $\lambda$ 和最大迭代数,初始化归属度矩阵 $A$ 和吸引度矩阵 $R$ 。根据式(1)计算基于欧氏距离的相似度矩阵,根据式(2)~(5)计算形态距离相似度矩阵,并根据改进熵权法计算不同相似度矩阵权重系数,根据式(14)完成对相似度矩阵的初始化。

**步骤 3 基于欧式形态距离的 AP 聚类结果。**更新吸引度矩阵和归属度矩阵,直至矩阵稳定或达到最

大迭代数  $t$ 。如果聚类数达到  $N_{\max}$ , 则进入步骤 4 中。否则, 调节  $p$ , 重复执行步骤 3。

**步骤 4** 确定最优聚类结果。根据聚类结果, 计算有效性指标, 选择  $I_{DB}$  最小时的聚类结果作为最优。

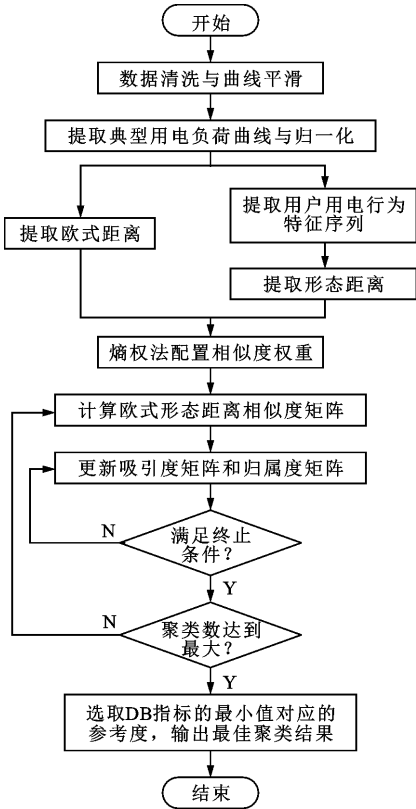


图 3 本文方法流程  
Fig. 3 Flow chart of proposed method

### 3 实验与分析

本文实验数据集包含两个部分。数据集 1 为 UCI 网站提供的一套标准合成时间序列数据集, 如图 4 所示。该数据集包含 600 个有标签时间序列数据, 分为 6 种不同的变化趋势, 每种变化趋势包含 100 条时序曲线, 使用数据集 1 验证本文方法的有效性。数据集 2 为爱尔兰居民用电实测数据。

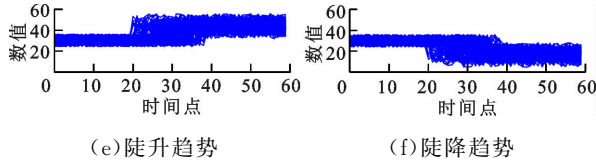
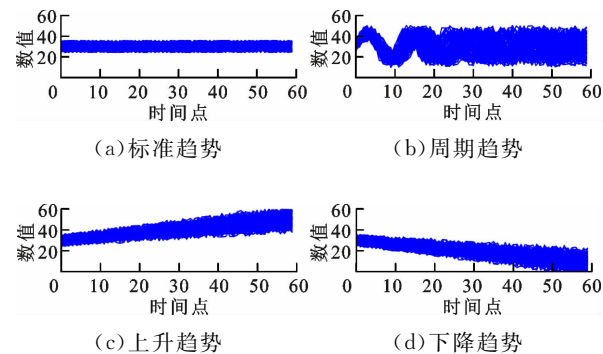


图 4 标准合成时间序列集  
Fig. 4 Standard synthetic control chart time series

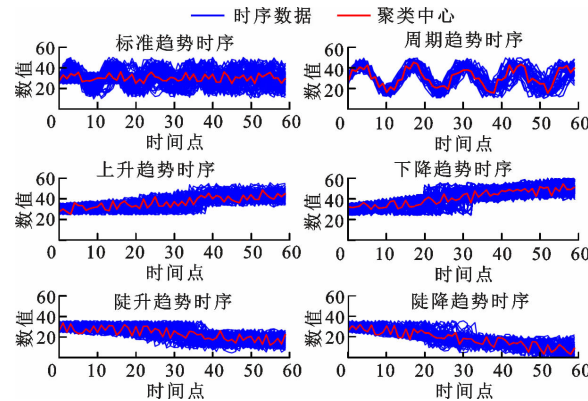
本文算例均在一台 Win10 操作系统 64 bit 的计算机上完成, 其配置为 Intel(R) Core(TM) i7-6700 CPU @3.4 GHz, 编程语言为 Python。

#### 3.1 形态距离有效性验证

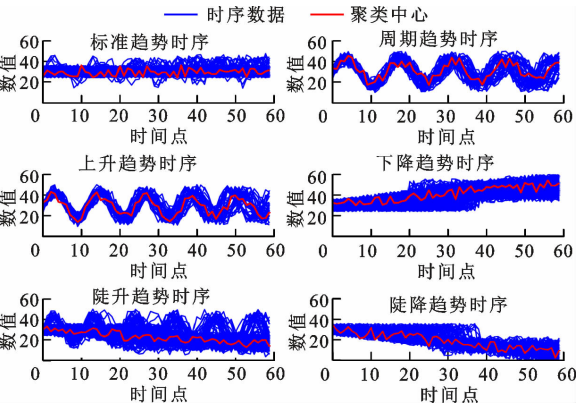
为了验证形态距离用于衡量负荷曲线相似度的有效性, 使用基于形态距离的 AP 聚类方法对数据集 1 进行聚类分析。由于数据集 1 时序曲线长度为 60, 设置阈值参数  $\epsilon$  为 6。AP 聚类方法通过参考度参数的设置来控制聚类数, 该参数一般设为相似度矩阵的最小值。本文设定  $p$  的初始值为矩阵元素最小值, 再通过调节该参数将各方法的聚类数调整为 6, 在此情况下对各方法聚类效果进行比较。

采用欧氏距离方法 (ED-AP)、形态相似距离方法 (MSD-AP)<sup>[21]</sup>、快速动态弯曲距离方法 (Fast-DTW-AP)<sup>[22]</sup>、动态弯曲距离方法 (DTW-AP)、形态距离方法 (MD-AP) 以及本文欧氏形态距离方法 (EMD-AP) 的聚类结果如图 5 所示, 实验中设定聚类数为 6。

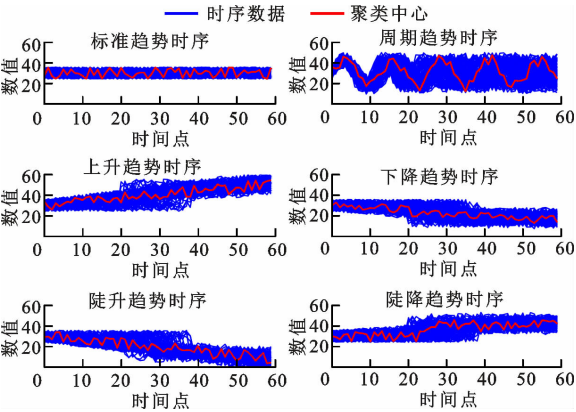
由图 5f 可知: 本文所提方法能够准确识别曲线整体的形态变化特征, 然而对于第一类振荡变化和第二类周期性变化的时序曲线辨别能力较差。正如在 2.1 小节中对形态距离的分析, 由于对原始用电数据进行了类属型转换, 所以原始曲线的数值差异性会被忽略。周期性曲线和振荡变化曲线的差异主要在于点对点之间的距离以及变化的周期性。形态距离主要针对于形态变化的特征, 忽略了点对点之



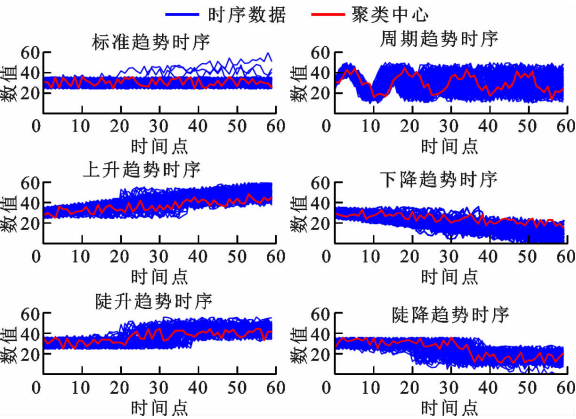
(a) 欧氏距离方法



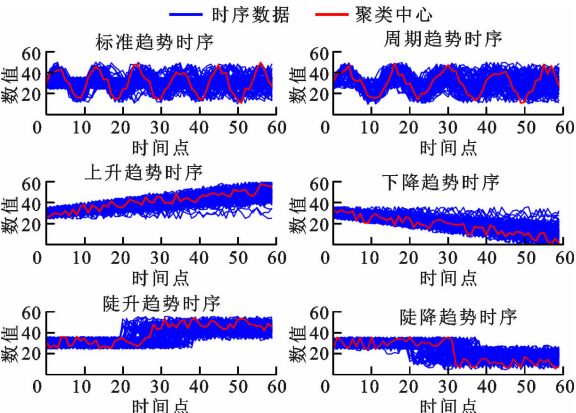
(b)形态相似距离方法



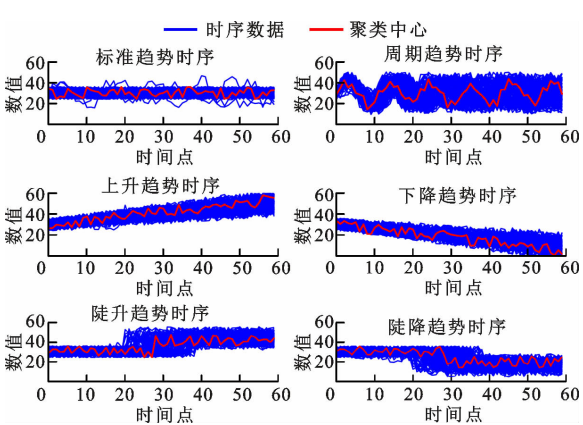
(c)快速动态弯曲距离方法



(d)动态弯曲距离方法



(e)形态距离方法



(f)本文欧式形态距离方法

图 5 不同相似度度量方法聚类结果对比

Fig. 5 Clustering results of different similarity measures

间数值的差异性,因此对于周期性特征和一般性波动特征的辨识度不高。对比图 5a、5b 可以看出,形态距离在形态趋势变化方面的识别效果明显优于采用欧式距离和形态相似距离等方法,后两种方法均无法分辨时间序列的渐变趋势和陡变趋势,而本文方法在周期性数据和标准数据的辨别方面逊色于 DTW。

为直观表示形态距离对数据集 1 的分类效果,本文给出了方法聚类前后簇的相同元素数。表 1 为使用形态距离对数据集 1 聚类的结果,表中, $V_i$  代表在聚类结果中同一类曲线在各类簇  $U_i$  的实际聚类情况,每一类曲线均为 100 条。

表 1 形态距离方法聚类结果

| Table 1 Results of morphological distance scheme |       |       |       |       |       |       |
|--------------------------------------------------|-------|-------|-------|-------|-------|-------|
| 聚类                                               | 曲线数   |       |       |       |       |       |
| 情况                                               | $U_1$ | $U_2$ | $U_3$ | $U_4$ | $U_5$ | $U_6$ |
| $V_1$                                            | 79    | 16    | 2     | 3     | 0     | 0     |
| $V_2$                                            | 41    | 59    | 0     | 0     | 0     | 0     |
| $V_3$                                            | 0     | 0     | 100   | 0     | 0     | 0     |
| $V_4$                                            | 0     | 0     | 0     | 98    | 0     | 2     |
| $V_5$                                            | 0     | 1     | 3     | 0     | 96    | 0     |
| $V_6$                                            | 0     | 0     | 0     | 8     | 0     | 92    |

由表 1 可知,形态距离方法对于数据集 1 中后 4 类变化特征明显的曲线辨别能力较强。考虑到用户实际用电负荷曲线时段性特征以及峰谷特征明显,并且在周期特征方面的识别问题也可结合其他方法进行优化,因此形态距离方法适用于负荷聚类的场景。

3.2 欧式形态距离度量方法聚类效果验证

欧式形态距离综合考虑了曲线动态波动特征和

整体分布特征,为检验该方法的有效性,将本文欧式形态距离的聚类结果与采用欧式距离、形态相似距离<sup>[21]</sup>、快速动态弯曲距离方法<sup>[22]</sup>、动态弯曲距离、形态距离的聚类结果就聚类质量及聚类效率进行综合比较。聚类质量通过外部评价指标  $I_{AR}$ 、 $I_{FM}$  以及内部评价指标  $I_{DB}$ 、 $I_{MDB}$  进行综合评价;聚类效率通过方法运行时相似度距离矩阵的计算时间  $T_{SCT}$ 、聚类时间和迭代数进行评估。为保证在数据集 1 上针对不同聚类方法聚类有效性指标均有统一的聚类质量评判标准,在 3.2 和 3.3 小节中聚类有效性内部评价指标  $I_{DB}$  和  $I_{MDB}$  的计算,选择各类簇的均值作为聚类中心,而在实测数据集上通过推举的聚类中心进行计算<sup>[22]</sup>。同时,提供数据集 1 的有效性指标帮助衡量聚类质量。

表 2 不同相似度度量方法聚类结果指标比较

Table 2 Clustering results of different similarity measures

| 方法                     | 外部评价指标   |          | 内部评价指标   |           | 计算效率        |         |             |     |
|------------------------|----------|----------|----------|-----------|-------------|---------|-------------|-----|
|                        | $I_{AR}$ | $I_{FM}$ | $I_{DB}$ | $I_{MDB}$ | $T_{SCT}/s$ | 聚类时间/s  | 总用时/s       | 迭代数 |
| ED-AP                  | 0.503 3  | 0.592 7  | 1.612 7  | 4.967 5   | 1.409 2     | 0.526 5 | 1.935 7     | 50  |
| MSD-AP <sup>[21]</sup> | 0.513 3  | 0.608 5  | 1.439 1  | 2.456 8   | 9.296 1     | 0.519 6 | 9.815 7     | 51  |
| FastDTW-AP             | 0.787 4  | 0.655 3  | 2.838 2  | 4.004 8   | 2 006.663 5 | 0.440 8 | 2 007.104 3 | 51  |
| DTW-AP                 | 0.851 2  | 0.875 8  | 2.485 2  | 2.149 4   | 1 725.165 6 | 0.470 1 | 1 725.635 7 | 43  |
| MD-AP                  | 0.810 3  | 0.807 1  | 2.963 1  | 2.789 7   | 51.267 9    | 0.502 1 | 51.770 0    | 44  |
| EMD-AP                 | 0.941 3  | 0.951 0  | 2.535 8  | 2.042 3   | 52.677 1    | 0.445 7 | 53.122 8    | 42  |
| 标准集                    | 1.000 0  | 1.000 0  | 2.521 5  | 2.071 6   |             |         |             |     |

在加入曲线整体形态变化特征后,方法收敛性增强,迭代数有所减少。相比于基于 FastDTW 和基于 DTW 等两种基于动态弯曲距离的方法,本文方法在更低的时间开销下能实现更优聚类质量,并且得到了更高的分类准确率以及聚类紧凑度。由图 5f 可知:欧式形态距离方法能有效辨别这 6 类不同变化特征的曲线,6 个类簇的分布情况非常接近标准集。对比 FastDTW 方法和 DTW 方法,欧式形态距离方法在聚类有效性和计算效率上均优于 DTW 方法,这也说明了采用阈值参数降低方法时间复杂度的有效性。

3.3 不同聚类方法聚类效果对比

采用基于 DTW 的  $K$  均值方法(DTW-KM)<sup>[13]</sup>、基于遗传算法的谱聚类方法(GA-NJW)<sup>[23]</sup>、 $K$ -medoids 方法<sup>[24]</sup>以及凝聚层次聚类方法<sup>[24]</sup>进行对比。为避免数据预处理操作对聚类效果的影响,分别对这些方法在原数据集与高斯平滑滤波处理后的数据集上进行测试,选取两者效果更好的结果。实验中

表 2 为不同相似度度量方法聚类结果比较,可以得出结论如下。形态距离方法与传统的相似度度量方法相比,在聚类准确度和精度方面都有一定的提升,而相似度矩阵计算时间也在可接受范围内,并且通过牺牲空间复杂度的情况下还可以对计算时间进行进一步优化。与基于 FastDTW 的计算方法和基于 DTW 的计算方法相比,形态距离方法大多数指标均优于 FastDTW 方法,在  $I_{AR}$  上和 DTW 方法有 4%左右的精度差距,同时  $I_{DB}$  也高于 DTW 方法,这也符合 3.1 小节中针对实际聚类结果得出的结论。但是,形态距离在  $I_{MDB}$  和  $T_{SCT}$  方面明显优于 DTW 方法的,具有较低的时间开销。欧式形态距离方法在  $I_{AR}$  上相对于欧氏距离方法和形态距离方法分别提高了 43.8%、13.1%,同时  $I_{DB}$  与标准集仅相差 0.014 3。

各方法聚类数均设为 6,关键步骤或模型参数如下:在 DTW-KM 方法中,DTW 的时间窗口长度与本文 LCS 方法一致,均设为 6, $K$  均值方法最大迭代次数为 300;GA-NJW 方法引用了原文的实验结果以作对比,并实现了文献中提出的 NJW( $M=11$ )方法;在  $K$ -medoids 方法<sup>[24]</sup>中,本文分别设置最大迭代数为 300、500 和 700 进行实验,并最终设置最大迭代数为 500。所有方法均完成了 20 次重复试验,取其最优。图 6 为部分聚类方法的聚类结果。

由图 6 可知: $K$ -medoids 方法对于前两类曲线的识别效果较差;层次聚类方法对于渐变趋势和陡降趋势的曲线识别效果较差;NJW 和 DTW-KM 方法对于陡升趋势和上升趋势的数据识别效果较差。

表 3 为多种聚类方法聚类结果的比较结果。可以看出,本文方法在  $I_{FM}$  和  $I_{AR}$  上均具有最优的结果,说明了欧式形态距离方法最终聚类结果的准确度更高。在  $I_{DB}$  上,欧式形态距离方法的聚类结果虽然大于  $K$ -medoids、层次聚类以及 NJW 方法的,但

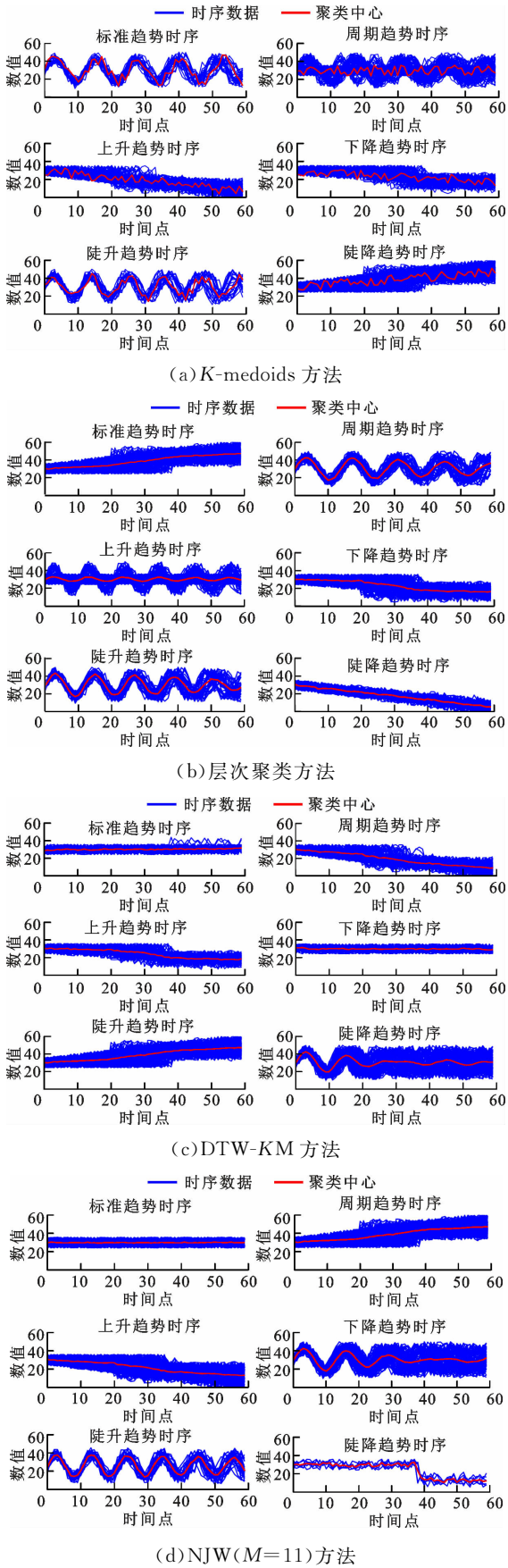


图 6 不同聚类方法聚类结果对比

Fig. 6 Clustering results of different clustering methods

是通过对比发现各类簇的分布情况更接近于原始数据集的。 $I_{MDB}$ 也具有相同的特点。因此,从多种方法的聚类结果进行分析,欧式形态聚类方法也具有较高的分类精度。

表 3 各聚类方法指标对比

Table 3 Clustering indexes comparison of different methods

| 方法                        | $I_{FM}$ | $I_{AR}$ | $I_{DB}$ | $I_{MDB}$ |
|---------------------------|----------|----------|----------|-----------|
| K-medoids <sup>[24]</sup> | 0.651 1  | 0.550 2  | 1.574 2  | 2.500 8   |
| 层次聚类                      | 0.690 2  | 0.606 5  | 1.363 8  | 2.334 7   |
| DTW-KM <sup>[13]</sup>    | 0.686 1  | 0.610 6  | 3.358 7  | 3.276 3   |
| NJW(M=11)                 | 0.716 5  | 0.636 2  | 2.185 8  | 1.904 5   |
| GA-NJW <sup>[23]</sup>    | 0.940 4  | 0.928 5  |          |           |
| EMD-AP                    | 0.951 0  | 0.944 9  | 2.535 8  | 2.042 3   |
| 标准集                       | 1.000 0  | 1.000 0  | 2.521 5  | 2.071 6   |

3.4 稳定性分析

在实际基于负荷建模的过程中,方法聚类结果的稳定性与可复现性也是非常关键的因素。20 次重复实验下的  $I_{FM}$  如图 7 所示。

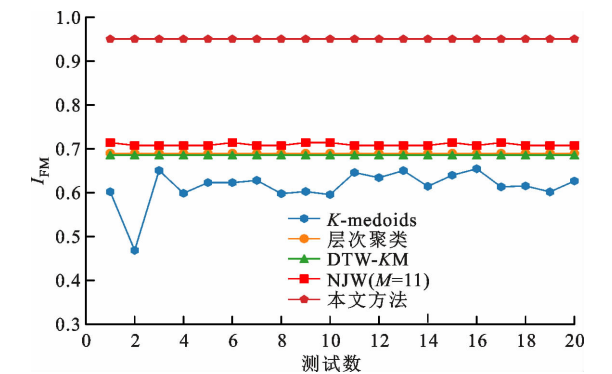


图 7 不同方法聚类稳定性对比

Fig. 7 Comparison of stability for different methods

由图 7 可知:K-medoids 方法 20 次聚类结果波动较大,且聚类质量较差;NJW 方法较 K-medoids 方法稳定性更好,但也存在陷入局部最优解的情况;DTW-KM、层次聚类方法以及本文方法在多次实验的情况下均能保持稳定的聚类结果,但本文方法聚类效果明显优于前二者,证明了本文方法聚类结果的优越性以及稳定性。

3.5 实际负荷数据集测试

本小节数据来自于 ISSDA 网站发布的爱尔兰智能电表实际量测数据<sup>[25]</sup>。从中随机选取 800 名用户为实验对象,实验结果如图 8 所示。

采用欧式形态距离的 AP 聚类方法对用户用电

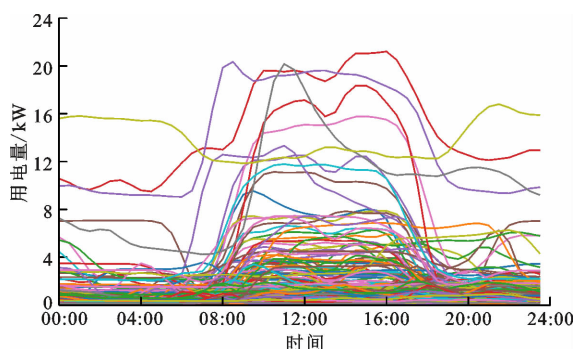


图8 800位电力用户用电实测数据

Fig. 8 Measured data of 800 power users

负荷曲线进行聚类测试,经熵权法计算,将式(6)中权重系数 $\alpha$ 和 $\beta$ 分别设置为0.501 5和0.498 5,阻尼系数为0.9。一般认为一个电力数据集的合理聚类数为2到 $\sqrt{0.5N}$ ,因此在本小节实验中设定聚类数取值范围为 $[2, 20]$ ,通过调节参考度 $p$ 获取对应聚类数,以 $I_{DB}$ 最小时对应聚类数作为最佳聚类数。图9为聚类数与聚类指标的变化关系。

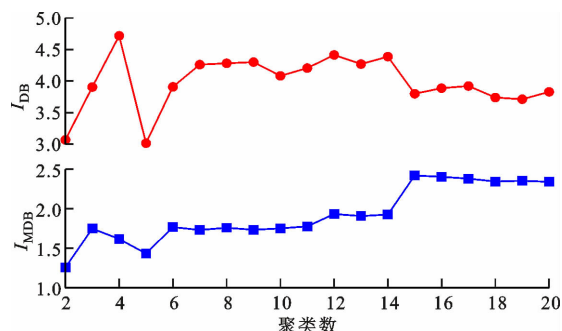


图9 聚类数与聚类有效性指标的变化关系

Fig. 9 Relationship between cluster number and cluster validity indexes

由图9可知:聚类数为5时, $I_{DB}$ 明显下降并且出现最小值,因此选择最优聚类数为5; $I_{MDB}$ 在聚类数为5时也出现极小值点,这也反映了将最优聚类数定为5的准确性和合理性。图10为聚类后的各类用户用电负荷曲线。

由图10可知:不同类簇之间用电负荷曲线的形态差异较大,用户的用电特征明显;A类用户为典型的峰平型负荷,此类用户用电特征明显,08:00—18:00为工作时段,在工作时段内具有较高的用电负荷,在12:00因为午休用电负荷有所下降,在非工作时段用电负荷骤降,这类用户可能为商业、工厂或办公楼等工作场所;B类用户在06:00—21:00出现用电高峰,其他时段用电量偏低,属于避峰型用户;C类用户在12:00和21:00出现用电高峰,属于双峰

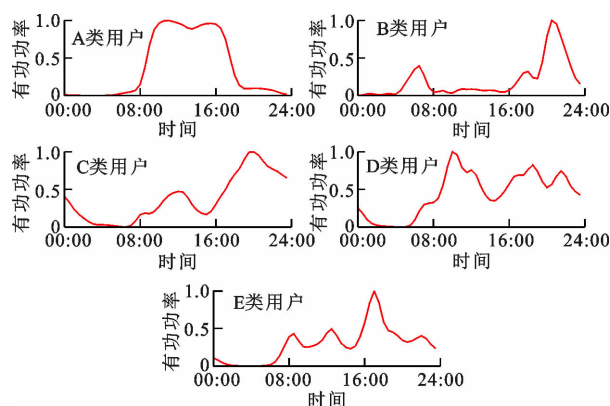


图10 电网实测数据聚类中心图

Fig. 10 Clustering center of measured data in power systems

型用电用户;D类用户分别在10:00、19:00和21:00有多次用电峰值出现,属于典型的三峰或多峰用户;E类用户在平时段用电平稳,而在18:00出现用电高峰,可能属于上班族,白天在外工作,18:00后回家,出现用电量峰值。

经本节分析可知,本文方法综合考虑了负荷曲线欧氏距离及形态距离,获得了考虑用能特性差异的归类结果,该归类结果说明了用户各种用电模式的分类情况,可用于为电力企业提供用户用电模式信息、电力负荷预测<sup>[26]</sup>以及制定能源规划策略等。同时,通过控制方法的阈值参数可以获取更精细化的分类结果,帮助电力企业推出个性化服务。

## 4 结 论

考虑到传统方法对于具有位移变化负荷曲线识别效果较差,本文提出了一种基于最长公共子序列的电力负荷曲线聚类方法,以模式匹配距离作为用户用电模式的差异度,将传统点对点距离计算简化为整数间的判等操作,有效降低了时间开销。相比于DTW方法,本文方法将时间开销从千秒级降低到了十秒级别,并且保留了对于陡变趋势以及渐变趋势这类相似变化特征的识别效果。考虑到方法在周期性曲线的识别方面具有一定的劣势,本文结合欧氏距离构造了一种兼顾负荷曲线的整体分布特征及局部动态变化特征的相似性度量——欧氏形态距离。算例证明:欧式形态距离方法具有较高的聚类精度,与多种经典方法相比,在聚类精度及聚类效率上均具有一定的优势,在实测电力数据集上提取的各类用电负荷曲线形态变化特征明显,可用于负荷预测、典型用电负荷提取、用户个性化服务及需求响应等领域。但是,本文所使用的AP聚类方法无法自动确定最优聚类数的遗留问题,有待在未来工作

中进一步改善。

## 参考文献:

- [1] 中共中央, 国务院. 关于进一步深化电力体制改革的若干意见: 中发〔2015〕9号 [A]. (2015-03-15)[2021-04-01]. <https://news.ncepu.edu.cn/xyd/llxx/52826.htm>.
- [2] 苏适, 李康平, 严玉廷, 等. 基于密度空间聚类和引力搜索算法的居民负荷用电模式分类模型 [J]. 电力自动化设备, 2018, 38(1): 129-136.  
SU Shi, LI Kangping, YAN Yuting, et al. Classification model of residential power consumption mode based on DBSCAN and gravitational search algorithm [J]. Electric Power Automation Equipment, 2018, 38(1), 38: 129-136.
- [3] ZHAO Qi, LI Haolin, WANG Xinying, et al. Analysis of users' electricity consumption behavior based on ensemble clustering [J]. Global Energy Interconnection, 2019, 2(6): 479-488.
- [4] 赵晋泉, 夏雪, 刘子文, 等. 电力用户用电特征选择与行为画像 [J]. 电网技术, 2020, 44(9): 3488-3496.  
ZHAO Jinquan, XIA Xue, LIU Ziwen, et al. User electricity consumption feature selection and behavioral portrait [J]. Power System Technology, 2020, 44(9): 3488-3496.
- [5] 张江林, 张亚超, 洪居华, 等. 基于离散小波变换和模糊  $K$ -modes 的负荷聚类算法 [J]. 电力自动化设备, 2019, 39(2): 100-106, 122.  
ZHANG Jianglin, ZHANG Yachao, HONG Juhua, et al. A load clustering algorithm based on discrete wavelet transform and fuzzy  $K$ -modes [J]. Electric Power Automation Equipment, 2019, 39(2): 100-106, 122.
- [6] 顾洁, 孟璐, 郑睿程, 等. 考虑集群辨识的海量用户负荷分层概率预测 [J]. 电力系统自动化, 2021, 45(5): 71-78.  
GU Jie, MENG Lu, ZHENG Ruicheng, et al. Load-stratified probability forecasting for massive users considering cluster identification [J]. Automation of Electric Power Systems, 2021, 45(5): 71-78.
- [7] 林顺富, 田二伟, 符杨, 等. 基于信息熵分段聚合近似和谐聚类的负荷分类方法 [J]. 中国电机工程学报, 2017, 37(8): 2242-2253.  
LIN Shunfu, TIAN Erwei, FU Yang, et al. Power load classification method based on information entropy piecewise aggregate approximation and spectral clustering [J]. Proceedings of the CSEE, 2017, 37(8): 2242-2253.
- [8] 黄鹤, 李昕芮, 吴琨, 等. 引入改进飞蛾扑火的  $K$  均值交叉迭代聚类算法 [J]. 西安交通大学学报, 2020, 54(9): 32-39.  
HUANG He, LI Xinrui, WU Kun, et al. Hybrid iterative  $K$ -means clustering with improved moth-flame optimization [J]. Journal of Xi'an Jiaotong University, 2020, 54(9): 32-39.
- [9] 卜凡鹏, 陈俊艺, 张琪祁, 等. 一种基于双层迭代聚类分析的负荷模式可控精细化识别方法 [J]. 电网技术, 2018, 42(3): 903-913.  
BU Fanpeng, CHEN Junyi, ZHANG Qiqi, et al. A controllable refined recognition method of electrical load pattern based on bilayer iterative clustering analysis [J]. Power System Technology, 2018, 42(3): 903-913.
- [10] 徐胜蓝, 司曹明哲, 万灿, 等. 考虑双尺度相似性的负荷曲线集成谱聚类算法 [J]. 电力系统自动化, 2020, 44(22): 152-160.  
XU Shenglan, SI Caomingzhe, WAN Can, et al. Ensemble spectral clustering algorithm for load profiles considering dual-scale similarities [J]. Automation of Electric Power Systems, 2020, 44(22): 152-160.
- [11] 冯志颖, 唐文虎, 吴青华, 等. 考虑负荷纵向随机性的用户用电行为聚类方法 [J]. 电力自动化设备, 2018, 38(9): 39-44, 53.  
FENG Zhiying, TANG Wenhua, WU Qinghua, et al. Users' consumption behavior clustering method considering longitudinal randomness of load [J]. Electric Power Automation Equipment, 2018, 38(9): 39-44, 53.
- [12] 李阳, 刘友波, 刘俊勇, 等. 基于形态距离的日负荷数据自适应稳健聚类算法 [J]. 中国电机工程学报, 2019, 39(12): 3409-3420.  
LI Yang, LIU Youbo, LIU Junyong, et al. Self-adaptive and robust clustering algorithm for daily load profiles based on morphological distance [J]. Proceedings of the CSEE, 2019, 39(12): 3409-3420.
- [13] TEERARATKUL T, O'NEILL D, LALL S. Shape-based approach to household electric load curve clustering and prediction [J]. IEEE Transactions on Smart Grid, 2018, 9(5): 5196-5206.
- [14] 宋军英, 崔益伟, 李欣然, 等. 改进分段线性表示与动态时间弯曲相结合的负荷曲线聚类方法 [J]. 电力系统自动化, 2021, 45(2): 89-96.  
SONG Junying, CUI Yiwei, LI Xinran, et al. Load curve clustering algorithm combining improved piecewise linear representation and dynamic time warping [J]. Automation of Electric Power Systems, 2021, 45(2): 89-96.

- [15] 林泽东, 鲁法明, 段华. 基于序列移动距离的用户行为挖掘与相似度计算 [J]. 计算机集成制造系统, 2017, 23(5): 1060-1068.  
LIN Zedong, LU Faming, DUAN Hua. User behavior mining and similarity computation based on sequence mover's distance [J]. Computer Integrated Manufacturing Systems, 2017, 23(5): 1060-1068.
- [16] 王潇笛, 刘俊勇, 刘友波, 等. 采用自适应分段聚合近似的典型负荷曲线形态聚类算法 [J]. 电力系统自动化, 2019, 43(1): 110-118.  
WANG Xiaodi, LIU Junyong, LIU Youbo, et al. Shape clustering algorithm of typical load curves based on adaptive piecewise aggregate approximation [J]. Automation of Electric Power Systems, 2019, 43(1): 110-118.
- [17] 葛乐, 陆文伟, 周志成, 等. 基于改进熵权法和灰色关联分析的变压器故障诊断 [J]. 电测与仪表, 2016, 53(12): 46-51.  
GE Le, LU Wenwei, ZHOU Zhicheng, et al. Fault diagnosis of transformer based on improved entropy method and grey correlation analysis [J]. Electrical Measurement & Instrumentation, 2016, 53(12): 46-51.
- [18] 张斌, 庄池杰, 胡军, 等. 结合降维技术的电力负荷曲线集成聚类算法 [J]. 中国电机工程学报, 2015, 35(15): 3741-3749.  
ZHANG Bin, ZHUANG Chijie, HU Jun, et al. Ensemble clustering algorithm combined with dimension reduction techniques for lower load profiles [J]. Proceedings of the CSEE, 2015, 35(15): 3741-3749.
- [19] 褚壮壮, 梁军, 张旭, 等. 基于仿射传播聚类算法的广义负荷稳态特性建模及其应用 [J]. 电力自动化设备, 2016, 36(3): 115-123.  
CHU Zhuangzhuang, LIANG Jun, ZHANG Xu, et al. Modeling of generalized load steady-state characteristics based on affinity propagation clustering algorithm and its application [J]. Electric Power Automation Equipment, 2016, 36(3): 115-123.
- [20] 叶林, 李稼辰, 路朋, 等. 基于近邻传播聚类与 MC-MC 算法的风电时序数据聚合方法 [J]. 中国电机工程学报, 2020, 40(12): 3744-3754.  
YE Lin, LI Jiachen, LU Peng, et al. Wind power time series aggregation approach based on affinity propagation clustering and MCMC algorithm [J]. Proceedings of the CSEE, 2020, 40(12): 3744-3754.
- [21] ZHANG Jinbao, ZHAO Yongqiang, KONG Lingxian, et al. Morphology similarity distance for bearing fault diagnosis based on multi-scale permutation entropy [J]. Journal of Harbin Institute of Technology (New Series), 2020, 27(1): 1-9.
- [22] JIN Yu, BI Zhongqin. Power load curve clustering algorithm using fast dynamic time warping and affinity propagation [C]//Proceedings of the 2018 5th International Conference on Systems and Informatics (ICSAI). Piscataway, NJ, USA: IEEE, 2018: 1132-1137.
- [23] 丁明, 黄冯, 邹佳芯, 等. 改进谱聚类与遗传算法相结合的电力时序曲线聚类方法 [J]. 电力自动化设备, 2019, 39(2): 93-99, 114.  
DING Ming, HUANG Feng, ZOU Jiaxin, et al. Power time series curve clustering method combining improved spectral clustering and genetic algorithm [J]. Electric Power Automation Equipment, 2019, 39(2): 93-99, 114.
- [24] 彭显刚, 赖家文, 陈奕. 基于聚类分析的客户用电模式智能识别方法 [J]. 电力系统保护与控制, 2014, 42(19): 68-73.  
PENG Xiangang, LAI Jiawen, CHEN Yi. Application of clustering analysis in typical power consumption profile analysis [J]. Power System Protection and Control, 2014, 42(19): 68-73.
- [25] MANNION C. Smart metering project: commission for energy regulation (CER) Ireland [C]//Proceedings of the 2010 IET Seminar on Smart Metering: Delivering a Smart UK. London, UK: IET, 2010: 11431756.
- [26] 李杰, 王秀丽, 邵成成, 等. 地区电网负荷预测的灰色 Verhulst 与系统动力学组合模型 [J]. 西安交通大学学报, 2021, 55(1): 68-77.  
LI Jie, WANG Xiuli, SHAO Chengcheng, et al. Grey verhulst and system dynamics combined model for regional power grid load forecasting [J]. Journal of Xi'an Jiaotong University, 2021, 55(1): 68-77.

(编辑 陶晴)